2005 American Control Conference
June 8-10, 2005. Portland, OR, USA

Information Theoretic Fault Detection

Alok Joshi, Paul Deignan, Peter Meckl, Galen King, and Kristofer Jennings

Abstract— **In this paper we propose a novel method of fault detection based on a clustering algorithm developed in the information theoretic framework. A mathematical formulation for a multi-input multi-output (MIMO) system is developed to identify the most informative signals for the fault detection using mutual information (MI) as the measure of correlation among various measurements on the system. This is a model-independent approach for the fault detection. The effectiveness of the proposed method is successfully demonstrated by employing MI-based algorithm to isolate various faults in 16-cylinder diesel engine in the form of distinct clusters.**

I. INTRODUCTION

This paper demonstrates a direct method that combines an information theoretic approach with cluster analysis to identify a lower-dimensional input space that can be employed for fault detection without building a model. The proposed method requires fewer measurements to perform fault detection.

A wide range of engineering systems such as automobile electronic throttle systems, aircraft control systems, induction motor control systems, and chemical process control systems require early fault-detection so that corrective measures can be taken to avoid system failure [1–4]. Various classes of methods for fault detection and identification include physical redundancy methods, model-based analytical redundancy methods, and model-independent methods.

In physical redundancy methods, multiple sensors or instruments are used to monitor a particular signal. If a single sensor from a set sensors indicates abnormal behavior of signal, then the case is treated as a sensor failure rather than system failure. If all the sensors depart from the normal operations and exceed some predetermined threshold, then a system fault is flagged [5]. This method requires multiple identical sensors, and the cost of implementation is high.

On the other hand, model-based analytical redundancy methods use a mathematical description of the system to artificially reproduce some of the measurements to detect anomalies in the operating system. These methods include expert systems [6], artificial neural network-based detection [7], Bayesian network-based detection [8], observer and filter-based detection methods [9], and parity space-based techniques [10]. Because of their cost effectiveness, the analytical redundancy approaches are more popular than physical redundancy-based approaches. However, these approaches require detailed knowledge of the physics of the system. The physical behavior of the system can be described in terms of the mathematical models derived from basic laws of physics or identified using system identification techniques. For systems with a large number of inputs and outputs, it is not always possible to get a precise description. Even if we are able to model MIMO systems precisely, the order of the model can be too large to implement a fault detection algorithm online. In such a situation, model-independent approaches can be employed effectively to detect and identify various faults in the system.

A wide class of model-independent techniques such as factor analysis, cluster analysis, principal component analysis, independent component analysis, and sequential multiple hypothesis testing are used for the fault detection [11–13]. These approaches generally involve grouping the signals based on a certain similarity measure of interest. These groups can be used for dimension reduction to find interesting patterns in the data. For instance, we can form various groups from a set of signals so that the signals within a group possess similar information about a certain output of interest and the signals in different groups possess non-redundant information about the output. From these groupings, various combinations of sensors can be selected to form a reduced input space. This input space can be used for fault detection. The main challenge in this approach is to identify the number of clusters and decide a criterion to partition the input space.

A wavelet transform-based approach [14] that monitors energy levels of signal is another of many model-independent approaches. The inspection of frequency-domain characteristics of the signal can also lead to detection of abrupt changes in the system. However, energy-based methods generally require fixed sampling rate that is in accordance with the Nyquist sampling theorem. Unfortunately, the data under investigation is not sampled at equal time intervals. Also, the number of sensors that monitor the engine is too large to build a model. For these reasons energy-based methods are not suitable for our application.

Section II describes the basic definitions used to calculate the mutual information of single-input single-output (SISO) or multiple-input single-output (MISO) systems. In Section III we develop mathematical formulae to calculate MI for multi-input multi-output (MIMO) system. We also discuss some properties of MI of MIMO systems. Section IV describes the acquisition and pre-processing of data obtained on Cummins diesel engine. Formation of distinct clusters for each type of fault encountered for the engine is demonstrated using 2- and 3-dimensional input spaces identified by MI-based algorithm. Section V summarizes the results.

A. Joshi, P. Deignan, P. Meckl, and G. King are with the School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907-2088. aajoshi@purdue.edu

K. Jennings is with the Department of Statistics, College of Science, Purdue University, West Lafayette, IN 47907-2067.

0-7803-9098-9/05/$25.00 ©2005 AACC

II. Mutual Information: Definitions for Single-Input Single-Output (SISO) System

This section describes basic mathematical tools required for reducing the dimension of the input space. All the variables obtained from the sampled signals are treated as discrete random variables. Entropy H is a measure of uncertainty of a random variable and is described for a discrete random variable Y as

$$H(Y) = -\sum_{y \in Y} p(y)\ln(p(y)), \tag{1}$$

where $p(y)$ is probability of each occurrence of $y \in Y$. If the possibility of every occurrence of a random variable in an event space is equally likely, then the random variable possesses high entropy. Entropy is measured in nats if natural logarithm is used for the formulation. If X and Y are two random variables then the entropy of Y with occurrence of X, i.e., the conditional entropy of Y with respect to X is defined as

$$H(Y \mid X) = -\sum_{y \in Y}\sum_{x \in X} p(x,y)\ln\left[\frac{p(x,y)}{p(x)}\right], \tag{2}$$

where $p(\cdot)$ represents probability of each outcome in X, Y or joint X–Y domain. Conditional entropy is the amount of uncertainty remaining in an output after the observation of a certain input variable or variables. The mutual information $I(X;Y)$ between the output random variable Y and the input random variable X can be given by

$$I(Y;X) = \sum_{y \in Y}\sum_{x \in X} p(x,y)\ln\left[\frac{p(x,y)}{p(x)p(y)}\right]. \tag{3}$$

Mutual information is a measure of reduction in the uncertainty of output Y when X occurs. Hence the alternate representation of (3) can be given in the entropy form by

$$I(Y;X) = H(Y) - H(Y \mid X). \tag{4}$$

The unit of mutual information is nats if the entropy calculation is based on the natural logarithm. The idea can be extended to a single output Y and multiple inputs where X becomes a vector of input random variables.

Details on information theory can be found in [15] [16]. The equations for mutual information or entropy require the calculation of joint-probability density functions. More information about calculation of joint-pdfs and criteria for selecting optimal bin number for each discrete variable can be obtained from [17].

III. Mutual Information: Mathematical Formulation for MIMO Systems

In this section we develop mathematical formulae for mutual information of multi-input multi-output (MIMO) systems and discuss their properties along similar lines to the formulation for SISO systems. The main advantage of this treatment is that it can be used to identify multiple sensor measurements with the maximum information about multiple faults occurring in the system. The algorithm based on the MIMO approach for identifying the most informative sensor measurements is faster than the algorithm for multi-input single-output (MISO) approach that is applied separately for each output of interest.

A. Mutual Information for MIMO System

Let $X_1, X_2, \ldots X_n$ be the discrete random variables forming a set of n inputs, $Y_1, Y_2, \ldots Y_m$ be the discrete random variables forming a set of m outputs, $p(\cdot)$ be the probability mass function i.e. $p(x) = P(\{X = x\})$, and A, B be sets of all possible values of x and y, respectively. From (3)

$$I(X_1,..X_n;Y_1,..Y_m) = \sum_{x_1 \in A_1} .. \sum_{y_m \in B_m} p(x_1,..x_n,y_1,..y_m)\ln\frac{p(x_1,..x_n,y_1,..y_m)}{p(x_1,..x_n)p(y_1,..y_m)} \tag{5}$$

Using (1-4) and the basic rules of probability, the formulation for the MISO system can be extended to develop a formula to calculate mutual information for MIMO system. The MI between n-dimensional random input vector and m-dimensional random output vector can be given as

$$I(X_1,..X_n;Y_1,..Y_m) = H(X_1,..X_n) - H(X_1,..X_n|Y_1,..Y_m) \tag{6}$$

B. Properties of MI for the MIMO System

The entropy of the input conditioned on the output is not the same as the entropy of the output conditioned on the input. The entropy of the input subtracted from the entropy of the input conditioned on the output is same as the entropy of the output subtracted from the entropy of the output conditioned on the input,

$$\begin{aligned} I(X_1,..\,X_n,Y_1,..\,Y_m) &= H(X_1,..\,X_n) - H(X_1,..\,X_n|Y_m,..\,Y_1) \\ &= H(Y_1,..\,Y_m) - H(Y_1,..\,Y_m|X_n,..\,X_1) \end{aligned} \tag{7}$$

Figs. 1 [a–c] represent idea of (6) using set operations when mutual information for the MIMO system is obtained from entropy of the input conditioned on the output. Figs. 1 [d–f] are graphical representation for the calculation of the mutual information obtained from the entropy of the output conditioned on the input. From Figs. 1 [c] and [f] it can be seen that the mutual information calculated from entropy of the output conditioned on the input and entropy of the input conditioned on the output is the same.

C. Software for Implementation

Equation (6) was computed using multidimensional histograms. The total number of dimensions of this histogram depend on the number of inputs and outputs considered for the analysis. The multidimensional histogram is generally of large size and requires large memory for storage. For example, an eight dimensional histogram with ten bins per dimension has 10^8 cells. If we have a thousand data points, at most thousand cells out of 10^8 cells will be occupied.

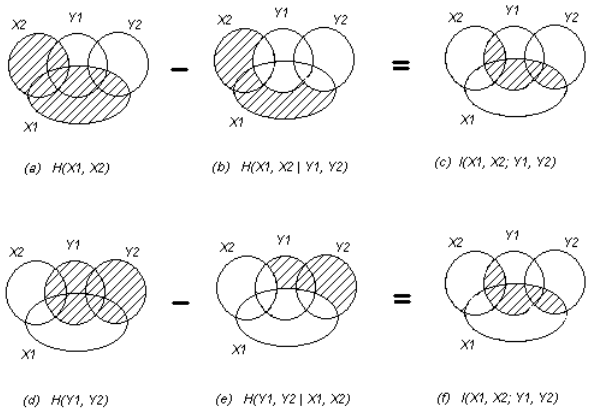

Fig. 1. Graphical representation for the calculation of mutual information for 2–input (X_1, X_2) 2–output (Y_1, Y_2) system

Instead of creating a multidimensional histogram, sparse matrix representation was used for the calculation of mutual information for a MIMO system. Programs were written in MATLAB where a preprocessed dataset, the number of outputs, and the number of inputs can be supplied as an argument to the function that calculates the mutual information.

IV. Application: Diesel Engine Fault Diagnosis

The proposed method for fault detection was tested on the data that were acquired on a Cummins QSK–60 unit. The single stage turbo QSK-60 engine is a 16-cylinder diesel engine with advertised hp of 2300 hp at 1900 rpm.

A. Data Acquisition and Pre–processing

Seventy-eight different measurements such as engine cylinder exhaust temperatures, oil cooler temperatures, input manifold temperatures, filter pressures, speed, and horsepower of the engine were monitored. Variables were named from u_1 to u_{78}. Table I gives a description of measurements used to monitor the engine. Various faults identified for the engine were documented over a period of time. These faults were mapped to the data acquired during the corresponding operation cycle to form a database consisting of 2135 data points and seventy-eight attributes. Data points [1–508] indicate a healthy state of the engine. Data points [509–821] indicate a state of the engine when a gasket failed in the flywheel housing. Data points [822–1443] and [1444–2135] indicate states of the engine when faults in the oil cooler and 8-R piston (8^{th} cylinder in the right bank) were identified, respectively. A binary column was added for each fault where 0 in the column indicates healthy state and 1 indicates the failed state. The first column corresponding to flywheel gasket failure consists of zeros [1–508] and ones [509–2135] indicating that the 508^{th} record separates the failed state from the healthy state of the engine. Similarly, the columns for oil cooler failure and piston scuff have all ones after the 821^{st} and 1443^{rd} points, respectively.

TABLE I

List of Measurements (Variables) on Cummins QSK–60 Type Diesel Engine.

LB- Left Bank, RB - Right Bank, F - Front, R - Rear
IMT - Input Manifold Temperature
EGT - Exhaust Gas Temperature

Variable	Description	Units
u_1	Brake Horse Power	bhp
u_2	ECM Temperature	^oF
u_3	% Acceleration Pedal	%
u_4	Instantaneous Engine Load	%
u_5	Oil Filter Differential Pressure	psig
u_6	Post Filter Oil Pressure	psig
u_7	Pre-filter Oil Pressure	psig
u_8	Oil Rifle Pressure	psig
u_9	ΔP Across Cooler $u_6 - u_8$	psig
u_{10}	LB Boost Pressure	psig
u_{11}	RB Boost Pressure	psig
u_{12}	LBF IMT	^oF
u_{13}	LBR IMT	^oF
u_{14}	RBF IMT	^oF
u_{15}	RBR IMT	^oF
u_{16}	Ambient Pressure	psia
u_{17}	Coolant Pressure	psig
u_{18}	Coolant Temperature	^oF
u_{19}	Rail Pressure	psig
u_{20}	Battery Voltage	V
u_{21}	LBR Compressor Inlet Temperature	^oF
$u_{22} - u_{37}$	Exhaust Temperature for 16-Cylinders	^oF
u_{38}	Avg. Exhaust Temperature	^oF
u_{39}	Engine oil Temperature	^oF
u_{40}	Engine Speed	rpm
u_{41}	Timing Pressure	psig
u_{42}	Fuel Temperature	^oF
u_{43}	Engine Condition Code	-
u_{44}	LB Avg Exhaust Temperature	^oF
u_{45}	RB Avg Exhaust Temperature	^oF
u_{46}	Custom	-
$u_{47} - u_{62}$	ΔT = $u_i - u_{38}$ for i = 22 to 37	^oF
$u_{63} - u_{78}$	20-Point Moving Avg for $u_{47} - u_{62}$	^oF

All signals were sampled with 10-Hz sampling rate. The acquired data was stored in a rolling buffer that ultimately saved every tenth point. Hence the effective sampling rate was 1 Hz. When the temperature and pressure sensor readings exceeded predetermined thresholds then a fault code was triggered. The buffer was traced back from that instant to 150 seconds before the event was triggered and then saved to the engine controller module (ECM) memory. Thus, 150 points were saved with effective sampling rate of 1-Hz when a fault code was triggered. The engine was assumed to be in the healthy state when all the sensor readings remained within predetermined thresholds. In the healthy state, data were saved every 30 minutes provided engine speed was within 1850-1925 rpm, percent acceleration and load were within 80-100 approximately. The main reason to adopt this sampling strategy was to reduce the volume of data that was being transmitted by all the running engines over a particular period of time. Unfortunately, this sampling setup resulted in a time-dependent sampling rate. In this scenario, frequency or energy-based fault detection methods may not give satisfactory results.

B. Off-line Data Analysis using Mutual Information

The main purpose of this analysis is to identify the variables reflecting significant change in the state of an engine when a particular type of fault occurs. The techniques described in Section III can be used to find an input space of the variables having maximum information regarding all the three types of failures. Taking into consideration the on-board implementation of models developed using the input-output space, we restricted the dimension of the input space to two variables at a time during the first phase of analysis. Though the reduced-order models are computationally efficient, care should be taken so that the effectiveness of the input space (in this case for fault detection and classification) is not lost by reducing the order of the input space. All the three binary columns, one for each fault, are considered the output of the interest. All possible 2-dimensional combinations were formed from the 78-dimensional input space. Mutual Information between the three outputs and every pair of variables from the 78-dimensional space was calculated.

Table II gives a partial list of 2-dimensional input spaces with high mutual information about all the faults, arranged in descending order of estimated mutual information. At this stage, the only rule to distinguish between informative input spaces and non-informative input spaces is to check if a particular input space achieves a performance specification for the fault detection. Also, the selection process for a particular input space from the entire set of highly informative input spaces will be influenced by the size and the cost of the candidate sensors. This procedure for input-space reduction does not use a process of forward selection or backward elimination of the sensors. Rather it cycles through all possible combinations of 2-dimensional input spaces that can be formed from the list of seventy-eight variables and lists them in descending order of the estimated MI values. The backward elimination or forward selection method may not find a global optimal set of sensors for a given dimension. For example, in a forward selection process u_{70}, (u_{70}, u_{76}), and (u_{70}, u_{76}, u_{78}) were identified as the most informative 1, 2, and 3-dimensional input spaces, respectively. From Tables II and III, it can be seen that these sets are not the sets with the highest estimated MI for a given dimension. Also, the backward elimination of sensors tends to not always identify the best informative pair for a given dimension. Hence we preferred exhaustive search of the sets of sensors for a given dimension.

C. Fault Detection and Classification

The MI between u_9 and all the outputs is 0.9965 nats. Also, the MI between (u_9, u_{74}) and all the outputs is 1.3069 nats. When u_{74} is added to the input space consisting of u_9, then the MI value between the input space and the output space jumps from 0.9965 to 1.3069 nats. Hence u_{74} adds the information about all the outputs that is not accounted by u_9. Thus, given a highly informative 2-dimensional input space the variables forming that input space add the information

TABLE II

LIST OF 2-DIMENSIONAL INPUT SPACES WITH HIGH MUTUAL INFORMATION WITH RESPECT TO ALL THE THREE FAULTS

No.	Input-Space	Mutual Information
1	(u_9, u_{74})	1.3069
2	(u_9, u_{76})	1.2895
3	(u_{70}, u_{76})	1.2774
4	(u_7, u_{70})	1.2670
5	(u_{73}, u_{78})	1.2633
6	(u_9, u_{52})	1.2629
7	(u_{74}, u_{78})	1.2595
8	(u_{70}, u_{71})	1.2542
9	(u_{75}, u_{78})	1.2335
10	(u_{71}, u_{74})	1.2316

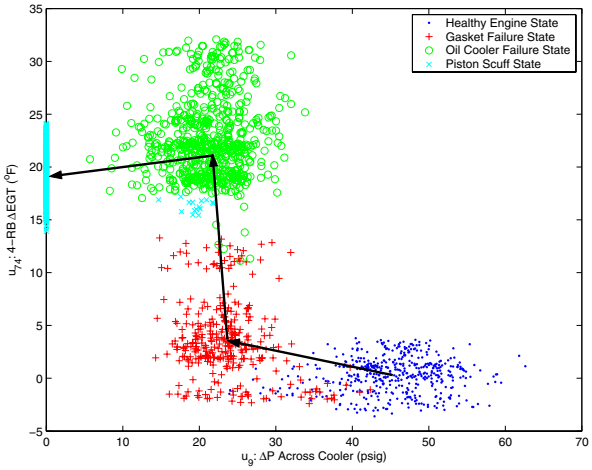

Fig. 2. The movement of a centroid during the change of a state using scatter plot of u_{74} : 4-RB ΔEGT Vs u_9 : ΔP across cooler

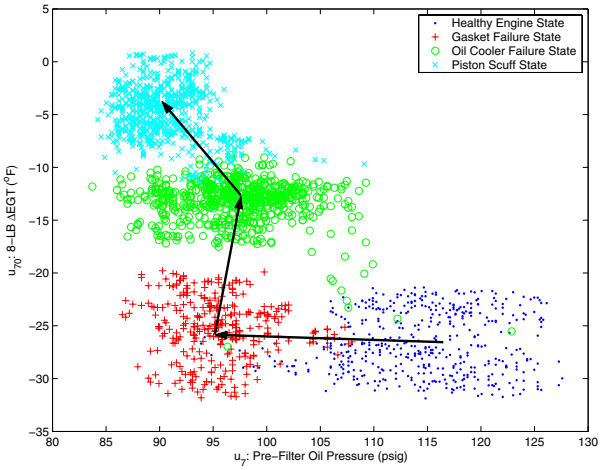

Fig. 3. The movement of a centroid during the change of a state using scatter plot of u_{70} : 8-LB ΔEGT Vs u_7 : pre-filter oil pressure

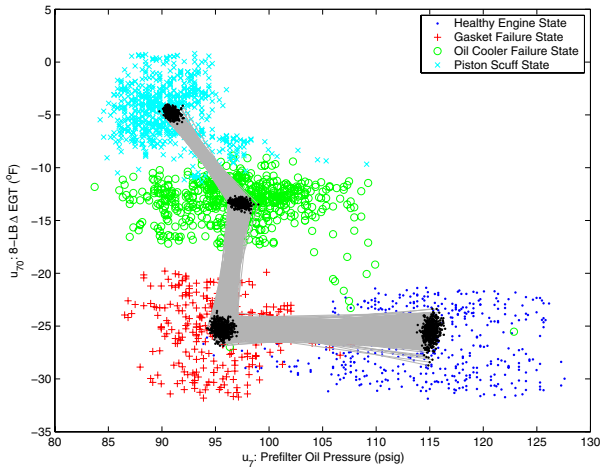

Fig. 4. Scatter plot of u_{70} : 8-LB ΔEGT Vs u_7 : pre-filter oil pressure. Bands from one cluster to another indicate the region where a displacement vector of the centroid will possibly orient when the state of the engine changes. Regions at the start and end of each band indicate possible location of the centroid when the engine is in a particular state.

about the output in non-redundant way. Fig. 2 shows scatter plot for (u_9, u_{74}) that is identified as the most informative 2-dimensional input space in Table II. Also, Fig. 3 shows the scatter plot for (u_7, u_{70}) that is one of the informative pairs identified in Table II. It can be observed that both the scatter plots reveal four distinct clusters, each representing a healthy state, a gasket failure state, oil cooler failure state, and a piston scuff state. This formation of distinct clusters in the scatter plot of variables forming highly informative input spaces can be used effectively for fault detection and classification.

The movement of the centroid of each cluster can be used for fault detection as the diesel engine goes from one state to another. The displacement vector of each centroid can be used as a measure of the fault detection, i.e., the distance as a measure of fault detection and the direction as a measure for classification. Figs. 2 and 3 also demonstrate the motion of a centroid for fault detection and classification using various input spaces identified in the previous subsection. Using a bank of moving average filters where number of filters is equal to the dimension of input space, the centroid of data acquired from the engine can be monitored.

The data under consideration were acquired from the unit that went through a specific sequence of the faults. To estimate confidence regions for the locations of centroid and their displacement vectors, 500 datasets were bootstrapped in blocks from the available time series and centroid locations were estimated in each simulation. These simulations were based on a dataset obtained from a single engine. Hence they may not capture variation in location of the centroid if another engine of the same type experiences a similar sequence of faults. Though we do not intend to use these confidence regions for on-board implementation, the procedure of resampling for the estimation will help us study the methodology that can be used with limited availability of similar datasets. Fig. 4 shows confidence regions for the location of centroid and corresponding displacement vectors estimated using these bootstrapped datasets.

TABLE III
LIST OF 3-DIMENSIONAL INPUT SPACES WITH HIGH MUTUAL INFORMATION WITH RESPECT TO ALL THE THREE FAULTS

No.	Input-Space	Mutual Information
1	(u_9, u_{74}, u_{76})	1.3425
2	(u_9, u_{76}, u_{78})	1.3418
3	(u_{70}, u_{76}, u_{78})	1.3391
4	(u_9, u_{38}, u_{74})	1.3375
5	(u_9, u_{62}, u_{76})	1.3369
6	(u_6, u_{67}, u_{78})	1.3368
7	(u_{62}, u_{70}, u_{76})	1.3365
8	(u_9, u_{71}, u_{74})	1.3362
9	(u_9, u_{48}, u_{76})	1.3359
10	(u_9, u_{19}, u_{74})	1.3358

Similar analysis can be done for a 3-dimensional input space. Table III gives a partial list of 3-dimensional input-spaces with high mutual information about all the faults, arranged in descending order of the mutual information. Figs. 5 and 6 demonstrate the formation of distinct clusters for each state of the engine using first and the third most informative input pairs shown in Table III. For the purpose of classification, output is coded as a categorical variable (1–healthy state, 2–gasket failure, 3–oil cooler failure, and 4–piston scuff state) and fit by linear regression using input space (u_9, u_{74}). 6% of total points were misclassified using this linear model. Total misclassification was 5.6% when a 3-dimensional input space (u_9, u_{74}, u_{76}) was used to fit a linear model. If the performance of 2 and 3-dimensional input spaces for detection of faults is compared based on misclassification using a linear decision boundary, then it can be concluded that a 2-dimensional input space is enough to form distinct clusters for four different states of an engine. The lowest possible dimension should be used if the performance of fault detection at that dimension is acceptable. A higher dimensional input space may be required if there are more faults to be detected. However, the Euclidean distance may not be a reliable metric for fault detection in higher dimensions. In that case, faults can be grouped so that the required dimension of the input space to detect these groups of faults can be restricted to two or three.

V. CONCLUSIONS

Formulae for the calculation of mutual information of a multi-input multi-output system was developed. Using a MI-based clustering algorithm, the formation of distinct clusters for each type of fault in the diesel engine was demonstrated successfully. Each cluster represents a different state of engine viz healthy state, a gasket failure state, oil cooler failure state, and piston scuff. The displacement

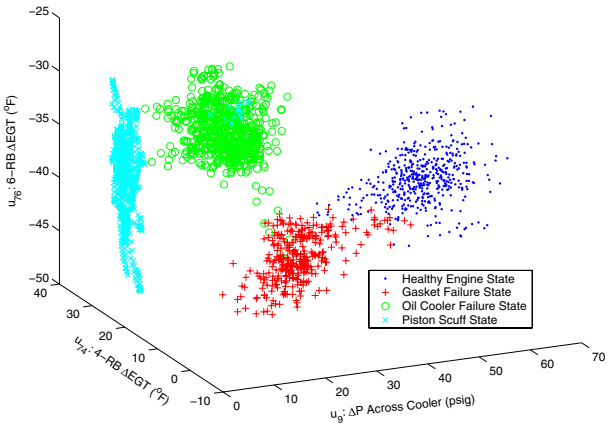

Fig. 5. Formation of distinct clusters using scatter plot of u_{76} : 6-RB ΔEGT Vs u_{74} : 4-RB ΔEGT Vs u_9 : ΔP across cooler

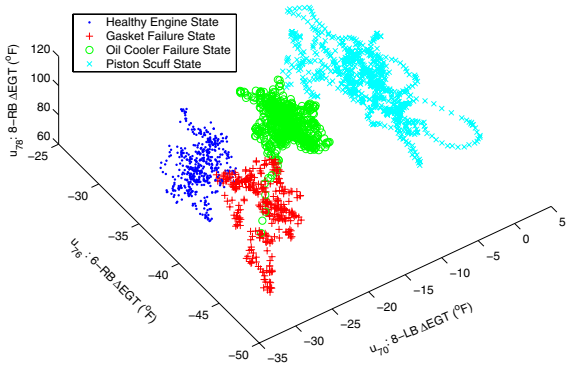

Fig. 6. Formation of distinct clusters using scatter plot of u_{70} : 8-LB ΔEGT Vs u_{76} : 6-RB ΔEGT Vs u_{78} : 8-RB ΔEGT

vector for the centroid of each cluster can be used for fault detection and isolation. The distance covered by the centroid can be a metric of fault detection. The direction in which the centroid moves can be used for isolation of the fault. Misclassification of 6% and 5.6% was observed when a linear classification model was built using the most informative two and three dimensional input spaces, respectively. Hence, two dimensional input space was sufficient to isolate the four states into distinct clusters. Using this reduced input space of sensor measurements and an equivalent number of moving average filters, the cluster centroid can be monitored for the on-line implementation.

The proposed method is model-independent and can be used effectively for highly non-linear systems. Since calculation of MI is based on the construction of a multi-dimensional histogram, data acquired in the transient as well as steady state of the engine with or without constant sampling rate can be used for the analysis.

VI. ACKNOWLEDGMENTS

We gratefully acknowledge the support of Cummins Inc. in making this research possible.

REFERENCES

[1] N. F. Thornhill, J. W. Cox, and M. A. Paulonis, Diagnosis of Plant-wide Oscillation through Data Driven Analysis and Process Understanding, *Control Engineering Practice*, 2003, Vol. 11, pp. 1481–1490.

[2] A. G. Parlos, K. Kim, and R. M. Bharadwaj, Sensorless Detection of Mechanical Faults in Electromechanical Systems, *Mechatronics*, 2004, Vol. 14, pp. 357–380.

[3] R. Conaster, J. Waganer, S. Ganta, and I. Walker, Diagnosis of Electronics Automotive Throttle Control System, *Control Engineering Practice*, 2004, Vol. 12, pp. 23–30.

[4] L. P. Myers, J. L. Baer-Riedhart, and M. D. Maxwell, "Fault Detection and Accomodation Testing on an F100 Engine in an F-15 Airplane," *NASA Technical Memorandum 86735,* 1985, pp. 1–12.

[5] Y. Liu, Y. Shen, and H. Hu, "A New Method for Sensor Fault Detection, Isolation and Accommodation," *Instrumentation and Measurement Technology Conference, IMTC/99. Proceedings of the* 16^{th} *IEEE*, 2004, Vol. 1, No. 24–26, pp. 488–492.

[6] R. Isermann and M. Ulieru, "Integrated Fault Detection & Diagnosis," *Proceedings of IEEE/SMC'93 Conference on Systems Engineering in the Service of Humans*, October 17-20, 1993, Le Touquet, France, Vol. 1, pp. 743–748.

[7] D. Capriglione, C. Liguori, C. Pianese , and A. Pietrosanto, On-line Sensor Fault Detection, Isolation, and Accommodation in Automotive Engines, *IEEE Transactions on Instrumentation and Measurement*, Aug. 2003, Vol. 52, No. 4, pp. 1182–1189.

[8] H. Kirsch and K. Kroshel, "Applying Bayesian Networks to Fault Diagnosis," *Proceedings of the* 3^{rd} *IEEE Conference on Control Applications*, 1994, Vol. 2, pp. 895–900.

[9] P. M. Frank and X. Ding, Survey of Robust Residual Generation and Evaluation Methods in Observer-based Fault Detection System, *Journal of Process Control*, 1997, Vol. 7, pp. 403–424.

[10] J. Gertler, Fault Detection and Isolation using Parity Relations, *Control Eng. Practice*, 1997, Vol. 5, No. 5, pp. 653–661.

[11] T. L. Lai, Sequential Multiple Hypothesis Testing and Efficient Fault Detection-Isolation in Stochastic Systems, *IEEE Transactions on Information Theory*, 2000, Vol. 46, No. 2, pp. 595–608.

[12] P. Huber, Projection Pursuit, *The Annals of Statistics*, 1985, Vol. 13, No. 2, pp. 435–475.

[13] T. Hastie, R. Tibshirani, and G. Walther, "Estimating the Number of Data Clusters via the Gap Statistics," ——, 2000, pp.1–19.

[14] J. Q. Zhang and Y. Yan, A Wavelet-based Approach to Abrupt Fault Detection and Diagnosis of Sensors, *IEEE Transactions on Instrumentation and Measurement*, October 2001, Vol. 50, No. 5, pp.1389–1396.

[15] C. B. Bell, Mutual Information and Maximal Correlation as Measure of Dependence, *Annals of Mathematical Statistics*, 1962, Vol. 33, pp 587–595.

[16] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley, New York; 1991.

[17] P. B. Deignan, G. B. King, and P. H. Meckl, "Combinatorial Optimization of Mutual Information Estimates Using Time-delayed Input Coordinates," *Intelligent Engineering Systems through Artificial Neural Networks*, November, 2003, Vol. 13, pp. 267-272.